

Evaluation of PATB
and Its Use for Selecting Personnel

STATINTL



January 1980

Evaluation of PATB and Its Use for Selecting Personnel

We were assigned the tasks of reviewing the evidence for the validity and reliability of PATB, its use in the Agency for selecting personnel for professional positions, its fairness for use with all applicants for positions and to make recommendations for improvements in the tests and their use.

Our assigned tasks did not include a study and evaluation of the intensive psychological assessment procedures. We have not obtained any data on these procedures and will make no comments about them.

We were also not assigned the task of reviewing and evaluating all of the procedures used to select personnel for Agency positions. We are aware that other procedures such as interviews, review of academic records, and recommendations are used in addition to or instead of PATB. Although we have not systematically studied these other procedures, we could find no evidence that they have ever been validated. The Uniform Guidelines on Employee Selection

~~ADMINISTRATIVE-INTERNAL USE ONLY~~

Procedures (1978) (^{1/}EE0C), apply not only to tests that are used to select personnel but also to all other procedures used for the same purpose.

In carrying out our assigned tasks, we discovered that a number of units in the Agency were using tests constructed in the Unit or taken from other sources to make decisions about employing candidates. We were not able to determine how extensive this practice is nor did we find any evidence that these tests had ever been validated. We think that we should call to the attention of responsible officials in the Agency the need to control the use of tests for selection and particularly to insist that no test be used for this purpose until it has been properly validated.

Nature of PATB and Its Use

PATB was constructed in the early or middle 1950's and was implemented around 1956 or 1957. It was designed for use with people who were applying for professional positions in the Agency. In this report we have focused on PATB Part I and Part II as it is currently used in the Agency.

^{1/} Equal Employment Opportunities Commission. Uniform guidelines on employment selection procedures. Federal Register, August 25, 1978, 43 (166) pp. 38296-7.

In carrying out our assigned tasks, we discovered that a number of units in the Agency were using tests constructed in the Unit or taken from other sources to make decisions about employing candidates. We were not able to determine how extensive this practice is nor did we find any evidence that these tests had ever been validated. We think that we should call to the attention of responsible officials in the Agency the need to control the use of tests for selection and particularly to insist that no test be used for this purpose until it has been properly validated.

Nature of PATB and Its Use

PATB was constructed in the early or middle 1950's and was implemented around 1956 or 1957. It was designed for use with people who were applying for professional positions in the Agency. In this report we have focused on PATB Part I and Part II as it is currently used in the Agency.

ADMINISTRATIVE INTERNAL USE ONLY

Nature of PATB

PATB is divided into 2 parts, one of which can be and is administered in field stations located in different geographical areas in the United States. The second part is administered by the staff of the Psychological Services Staff (PSS) in the Washington Office.

A short description of the content of PATB Part I and Part II is given below.

Part I

<u>Tests</u>	<u>Time limits</u>	<u>Short Description of Content</u>
Vocabulary Span	15 minutes	60 multiple-choice items in which a word is given and the examinee has to select a synonym.
Verbal Comprehension	25 minutes	The test contains 11 reading passages predominantly selected from literary essays and 41 multiple-choice items based on the

ADMINISTRATIVE INTERNAL USE ONLY

~~ADMINISTRATIVE INTERNAL USE ONLY~~

passages. However, only

31 items based on 9

reading passages are

scored.

Abstract
Reasoning

25 minutes

30 multiple-choice items
using figural symbols
(non-verbal) arranged in a
3X3 matrix form. The
examinee must deduce the
progression of changes
occurring in the matrix and
identify what the content
of the last cell in the
matrix would be.

Arithmetic
Reasoning

30 minutes

30 multiple-choice verbal
problems in arithmetic.

Language
Aptitude

Not Known

Test consists of a series of
tasks requiring the
examinee to learn an
artificial language. 59
multiple choice questions
are based on the learning
tasks.

~~ADMINISTRATIVE INTERNAL USE ONLY~~

Strong-Campbell Untimed
Interest
Inventory

325 items relating to
occupational titles,
school subjects, activities,
amusements, types of
people, activity preferences
and personal characteristics.
For the majority of the
items on the inventory,
the examinee marks like or
dislike or indifferent.
The inventory yields 158
different scores.

Part II

Interpretive Untimed
Reasoning

Four problem situations-
one a genetic chart, one a
report of an experiment
and two presenting graphical
material - are given
followed by a series of
statements. The examinee
is asked to judge the
degree of truth or
falsity of each statement.
There are 40 statements.

Biographical Information	Untimed	123 questions about personal background and experience and recreational preference. Many of the items duplicate material found in the Personal History Statement.
Preferences and Habit Survey	Untimed	108 questions to be answered yes, no or unde- cided. The results of this survey are referred to as temperament and are reported as 7 separate scores using the labels: quick, physical, outgoing, predominant, self-confident, solitary and question.
Work Environ- ment Inventory	Untimed	105 items to which examinee responds on a rating scale from 1, meaning highly desir- able, exactly what he/she would want, to 5, meaning highly undesirable, he/she would

probably refuse the job.

Scores on this instrument are reported as work attitudes. 14 different scores are extracted from the 105 items.

Numerical Operations	10 minutes	180 very elementary numerical computational problems.
Considerations	3 minutes per question	Examinee is given three situations: negotiating a trade treaty with Soviet Russia, selecting a site for a new plant of a manufacturing corporation, and an episode for a detective story. The examinee is required to write down as many questions or considerations that he/she can think of that are relevant. The score is the number of responses that the examinee gives. Quality of the responses

~~ADMINISTRATIVE INTERNAL USE ONLY~~

incorrect

or repetition do not enter
into the score.

Contemporary
Affairs Test

Unknown

A number of somewhat lengthy
reading passages are given
describing a current
event. The examinee is
required to identify the
leader or the foreign
country described. The
latter sometimes has to be
identified on a map.
There are 50 multiple
choice questions.

Essay

30 minutes

Examinees are given three
topics - At what point
does a job become a
career?, A significant
personal experience, and
The role of personalities
in world affairs. The
examinee selects one topic
and writes an essay on it.
No systematic guides are

~~ADMINISTRATIVE INTERNAL USE ONLY~~

ADMINISTRATIVE-INTERNAL USE ONLY

used to score the essays.
They appear to be scored
impressionistically and no
quantitative scores are
reported.

*hence, they
are not
scored*

We could find no rationale for the assignment of the tests to the different parts of the battery. Except for the Strong-Campbell Interest Inventory, all of the tests in Part I are cognitive or intellectual tests, but some of the tests in Part II, Interpretive Reasoning, Considerations, Numerical Operations, Contemporary Affairs Test, and the essay, are also cognitive tests. Therefore, the assignment to different parts of the battery does not appear to be based on the cognitive - non-cognitive dimension. Three of the tests in Part II, Considerations, Numerical Operations and essay, have to be hand-scored but all the others can be machine scored; so the division does not appear to be based on this factor. The division could not be based on the fact that the tests in Part I have more relevance for jobs in the Agency and higher validity for predicting job success. We could find no evidence that systematic job analyses have been done and, as we will show in a later section of the report, none of the tests either in Part I or Part II have

ADMINISTRATIVE-INTERNAL USE ONLY

~~ADMINISTRATIVE-INTERNAL USE ONLY~~

The authors are either ignorant, blind, or unprofessional.

been shown to have any consistent validity for predicting job success. We suspect that someone a long time ago decided that the battery should be divided into two parts so that each part could be administered in 3 or 4 hours; and, having made this decision, he then somewhat arbitrarily assigned tests to the two parts.

Why is this important? What is the reason for trying so "hard" to find a rationale for the items in Part 1 and those included in Part 2? The authors seem to be spinning their wheels on an irrelevant issue.

Norms for PATB were established by testing people who were employed in professional jobs in the Agency in the 1950's. Separate norms were developed for males and females. The males used to establish the norms had longer tenure with the Agency, higher G-S levels and more advanced education than did the females. We suspect, but cannot prove definitively, that the females used to norm the test in the 1950's are not representative of the females in the applicant group today. We also suspect that minorities were either under-^{no basis for this claim at all}represented or not represented at all in the normative groups.

Then it shouldn't be said at all.

The descriptions of the norm groups do not give information on race or ethnicity. This is a serious omission. The omission of these data indicate that the results of PATB cannot be interpreted and probably should not be used for

so how is it that the above conclusion is arrived at?

minorities. The tests need to be renormed using a current applicant group.

I have no objection to this, but I don't expect it to make PATB a better instrument.

Not available at the time of Norming, it was illegal to collect this info just a few short years ago.

Use of PATB in the Agency

The results from PATB, if they are used at all, enter into

~~ADMINISTRATIVE-INTERNAL USE ONLY~~

decisions about employment at a relatively late stage of the total selection process. As we understand the employment process, a typical sequence goes something like the following. The first contact that a person who is interested in employment in the Agency has is with a field recruiter. The field recruiter may interview the person either by telephone or in person. On the basis of the interview, the recruiter may decide that the person is not suitable for a job with the Agency and then does not give the person an application form. These people never take any part of PATB. The field recruiter may decide that the person does have potential for employment and gives the person an application form. Some of these people never complete or submit the application form and are never considered for employment. Again, test scores play no part. If the application form is completed and submitted, then persons outside of the Washington area are assigned to one of the field settings to take PATB I and those within the Washington area are assigned to a Headquarters office to take both PATB I and PATB II. Answer sheets from the field settings are sent to PSS and scored but no report of the test scores is prepared.

When the applicant's Personal History Statement is received at headquarters, the Skills Bank in OP// prepares a list of names of applicants called an Applicant File Listing

that includes the applicant's name, education and relevant experience, foreign language capabilities, acceptable salary level, recruiter's recommendation, and prior commitments made to the applicant. No test scores are reported on this form. The listing is circulated among the various units of the Agency. If no one indicates an interest in an applicant within 10 days from the date of listing, the applicant's file is sent to storage where, at the present time, it is nearly impossible to retrieve it. All such applicants are essentially lost as far as possible employment is concerned.

If a unit expresses interest in an applicant, the manager or his representative can request the file or can go to the Skills Bank to examine the applicant's file. Generally, the file, at this point, contains no record of test scores. If the manager is still interested after examining the file and wants to see the results of PATB, the Skills Bank sends a request to PSS to prepare a narrative report of test performance. The managers in the units never see actual test scores of applicants; they see only the narrative report. It is only at this point and for the managers requesting test scores that PATB could influence the decision to employ a person.

We could not determine precisely how many applicants for professional positions in the Agency are required to take

~~ADMINISTRATIVE-INTERNAL USE ONLY~~

PATB and, if they are required to take it, how many personnel managers in the units use the results of PATB to make employment decisions. The Director of PSS estimated that only about one-half of the applicants for jobs in the Agency take PATB. However, we are not sure whether estimate applies to applicants for any type of job in the Agency or only to professional jobs. In a survey done by the OIG Survey Team of EOD's from 1 October 1977 to August 1979, 63% of the EOD's to professional positions reported that they had taken PATB and 37% reported they had not. We could find no written policies as to who is or is not required to take the tests. Some exceptions, such as those for applicants for professional positions that required highly specialized knowledge or competencies that are not appraised by PATB or for applicants who were directly recruited because they were known to have expert knowledge in an area of high priority to the Agency seem reasonable. Nothing could be gained and much could be lost if these people were required to take PATB. However, it did not seem reasonable to us that, among candidates with similar educational and experience backgrounds applying for the same type and level of professional job, some were required to take PATB and some were not. We think, some guidelines should be developed for

I question this especially in view of job adjustment problems we have seen in some test types.

~~ADMINISTRATIVE-INTERNAL USE ONLY~~

~~ADMINISTRATIVE-INTERNAL USE ONLY~~

the PATB requirement and that such guidelines be followed by all units. If the decision to require or not to require PATB for professional jobs is made by individual managers,^{1/} there is a high potential for violating the EEOC guideline on disparate treatment of applicants.

The best information that we could find on the extent to which managers in the units use PATB results also came from the survey of supervisors of EOD's from 1 October 77 to August 79 done by the OIG Survey Team. Four hundred ninety-one supervisors from 44 components of the Agency responded to the questionnaire. Of these, 65% stated either that PATB was not administered to applicants or that they had no opinion about the usefulness of PATB or that PATB results were not very useful and tended to be disregarded. Eleven percent indicated either that they tend to ignore PATB results if most other factors are positive or that PATB results are used mostly to eliminate weaker applicants. Only 26% of the supervisors reported that an applicant

^{1/} Equal Employment Opportunity Commission. Uniform guidelines on employment selection procedures. Federal Register, August 25, 1978, 43 (166), p. 38300, Sec 11.

~~ADMINISTRATIVE-INTERNAL USE ONLY~~

~~ADMINISTRATIVE-INTERNAL USE ONLY~~

is rarely selected without positive PATB scores or that PATB scores are one of several major determinants in making selection decisions. These results suggest that scores on PATB do not have a major role in selection decisions for the Agency as a whole but do play a significant, sometimes a major, role in certain managers' decisions. However, one must remember that these are reports of what supervisors say that they do. How well what was said matches the reality of what is done is still unknown. *done*

The variations that exist in the requirement for taking PATB and in the use of the results have disastrous implications for validation research. Such variations reduce the number of employees who have test scores thereby reducing the size of samples that can be used to study the validity of PATB. They also introduce unknown biases in the employee samples and, probably, into the criterion ratings of job effectiveness.

Validity and Reliability of PATB

We have presented a detailed critique of the validity and reliability of PATB in Appendix 1. In this section we will summarize the major findings of the detailed critiques.

~~ADMINISTRATIVE-INTERNAL USE ONLY~~

The reader who is interested in determining the documentation for our findings and a more complete discussion of them should read Appendix 1.

Rationale for PATB

When one sets out to construct a battery of tests to select personnel for jobs, the first step in test construction is to do a systematic analysis of the jobs to determine what knowledges, skills and competencies are needed to perform the job. Tests would then be constructed or selected from existing sources to appraise these knowledges, skills and competencies. We searched for but could find no evidence that the construction of PATB was based on a systematic analysis of jobs in the Agency. We could not find any material that explained nor could anyone tell us why the tests that comprise the PATB were chosen. The absence of evidence on job analyses and on the rationale for choosing the tests in the battery casts serious doubt on the content validity of the battery. The lack of job analysis data violates the standards for selection tests set by APA^{1/} and EEOC.^{2/}

Not so

1/ Standards for Educational and Psychological Tests. Washington, D.C.: American Psychological Association, 1974 p46 Standard E12.4.

2/ Op. Cit. p. 38300, Sec. 14A.

~~ADMINISTRATIVE-INTERNAL USE ONLY~~

Once a battery of tests for selecting personnel has been constructed, several types of studies need to be done to clarify what the tests are appraising and what scores on the tests mean. For example, correlations of the tests in PATB with other tests of known validity should have been done. Such studies have not been done. Factor analytic studies of the battery should have been done to clarify the meaning of the test scores. The Chief of PSS told us that factor analytic studies had been done but were not kept, and evidently not used. The type of validity that we are discussing here is called construct validity. It is particularly important in tests or scales that purport to appraise aspects of personality and temperament. The work attitude and temperament scales on the PATB are examples of these kinds of tests. To date no evidence has been produced to show what these tests are appraising or what scores on them mean.

Not true!
AARP study for ODP,
IC vs WAIS study
SCIT, CPI

Evidence on Criterion-related Validity

In the absence of evidence of content and construct validity, we searched for evidence of criterion-related validity; i.e. evidence that performance on the tests

Ignoring the data.

~~ADMINISTRATIVE-INTERNAL USE ONLY~~

ADMINISTRATIVE-INTERNAL USE ONLY

and scales of PATB was significantly related to performance on the job. The ideal study to determine the criterion-related validity of a test or battery of tests requires that (1) the tests be administered to a large and representative group of applicants; (2) all, or a random sample, of the group tested be placed in the same or highly similar jobs; and (3) reliable and relevant appraisals of their job performance be subsequently obtained. In the Agency, as in any other practical setting, none of these requirements of an ideal study can be fully met.

But if they aren't available, they can't be used!

not possible in the Agency. Preselection to get only the best applicants makes this impossible. !!!

★★

The number of employees in the same or highly similar jobs in the Agency is small and the numbers available for validity studies are reduced because not all employees are required to take PATB. The persons in the applicant group who are employed by the Agency do not represent a random sample of applicants but instead are highly selected. Some of the selectivity has been direct; i.e., people have been selected because they have high test scores and some rejected because they have low test scores. Some of the selectivity has been indirect; i.e., it results from selecting people who have performed exceptionally well in college or have earned advanced degrees. When the less capable applicants

textbook approach - does not work in the real world applied world.

ADMINISTRATIVE-INTERNAL USE ONLY

ADMINISTRATIVE-INTERNAL USE ONLY

have been excluded from jobs, validation results on the remaining cases will be distorted and generally weakened. In the extreme case, if all employees had identical scores on a test, that test could not predict differences in performance even though the competency measured by the test is a critical requirement for successful performance on the job.

we approach this due to extensive pre-selection from transcripts interviews, etc.

In a number of the studies on criterion-related validity that we examined, the investigators invested considerable time and effort to obtain good ratings from supervisors of job performance. These ratings, called criteria, were not developed by systematic analyses of jobs or job performance but by discussing the jobs with the supervisors. As a result, serious doubts can be raised about the relevance and validity of the criteria as measures of job performance.

Experience with supervisor's ratings of job performance over the past 60 years has indicated that they tend to be unreliable and often biased by a variety of factors other than job performance. Realistically, though, these types of criteria are the only ones that the investigators of the validity of PATB could possibly obtain.

School teachers view! Ask any of the managers in these studies and they will tell you that the criterion measures derived are both valid and accurate.

Well, well!

Thank you!

One additional factor has also hindered the efforts to obtain not only criterion-validity data on PATB but also

ADMINISTRATIVE-INTERNAL USE ONLY

ADMINISTRATIVE-INTERNAL USE ONLY

other types of validity data. In order to do a validity study, a high level of cooperation must exist among the managers of the units where the employees are and the psychologists conducting the studies. Our discussions with the psychologists doing the validity studies indicated that

Not really true. We just don't have the charter to demand such a degree of cooperation, or to initiate studies on our own.

the necessary level of cooperation could not be obtained from many of the units in the Agency. They also indicated that there was no higher level of administration to which they could turn to secure the necessary cooperation.

also in the DDC the "need to know" comparison takes its get in the way either really over a national get

or the staff is IB slashing of Research Branch their reasons for calling the evidence "weak and unconvincing" are themselves "weak and unconvincing"

The reader of this report should keep the aforementioned problems in mind in reading the comments that follow.

Although we found the evidence for criterion-related validity to be weak and unconvincing, the reason for this state of affairs rests as much or more on limitations of the settings in which the studies were done as on the inadequacies of the investigators.

Wow!!

We have reviewed 23 studies that purported to present evidence on the validity of PATB. In general we found the quality of reporting in the studies to be very poor. The samples of subjects used in the studies were inadequately described, and, when they were described, only sex of the subjects was mentioned. The data generated in the study

The consumer of the report was the sponsoring office, not a psych journal reader-ship. !!!

B.S. Read them again, folks.

ADMINISTRATIVE-INTERNAL USE ONLY

~~ADMINISTRATIVE-INTERNAL USE ONLY~~

were usually incompletely reported and the conclusions reached were frequently not justified by the data presented in the study. We had to discard 6 studies because they reported no data, one study because the statistical analyses of the data were inappropriate, and 2 studies because they did not report any data on the PATB. After discarding these studies we were left with 14 studies, 4 of which were related to foreign language training and 10 to job performance in various units in the Agency. A total of 14 studies is rather scanty ^{value judgment} considering that PATB has been in operational use for more than 20 years which, to us, indicates the lack of any systematic plan for validating PATB. *Probably accurate - all previous studies in this field were not validated.*

We reviewed the 4 studies that reported data on the relationship between scores on PATB and success in foreign language training. Only 2 tests, reading vocabulary and reading comprehension, had consistent correlations with success in foreign language training and these were consistent only for foreign language training in French and Spanish. Scores on the language aptitude test, which one would expect to be related to foreign language training, *fine* did not show consistent relationships. In general, on the basis of the results of these 4 studies, one would have to conclude that the scores on PATB provide very little help in predicting success in foreign language training. — *Probably a fair statement. This may be anecdotal but since as far as I was told, the speed of learning was certainly related to high & low scores. Does John Ross want anything on this?*

~~ADMINISTRATIVE-INTERNAL USE ONLY~~

~~ADMINISTRATIVE-INTERNAL USE ONLY~~

On the basis of our review of the 10 studies that investigated the relationships among scores on PATB and criteria of job performance, our findings were as follows.

1. No consistent pattern of correlations for similar jobs in the Agency or for similar criteria of job performance has been found. In all of the studies, the number of subjects used has been small in relation to the number of predictor variables (test scores) used, and the number of significant correlations obtained did not exceed what one would expect by chance. The correlation data provide meager, if any, support for the criterion-related validity of PATB.

*Not always
the case.*

*Inconclusive
value judgement.*

2. The equations generated by multiple-regression analysis or discriminant analysis to predict job performance have been based on extremely small and inadequate samples of employees and have not been cross-validated. These equations should not be used to select or place personnel in Agency jobs until they are cross-validated.

*Incorrect.
see my AAP study.*

3. The samples of employees used to study the criterion-related validity of PATB have been composed solely or primarily of ^{*How do they know?*} white males. Females have been underrepresented or not represented at all in these

*Compared to
what?*

22
~~ADMINISTRATIVE-INTERNAL USE ONLY~~

ADMINISTRATIVE-INTERNAL USE ONLY

samples, and there is no evidence that minorities are represented in these samples; therefore, equations to predict job performance should not be used for females and minorities.

4. No validity studies have been done using the writing sample, and the Strong-Campbell Interest Inventory, and no validity data are available for the majority of items on the Biographical Information Inventory.

5. The evidence on validity presented in the

10 studies does not meet minimum standards for validity

1974 → 1/ set by APA or EEOC guidelines. 2/ ← These are the 1978 guidelines

again, the studies were done for office mgmt not journal editors.

Given the meager, unreliable and unstable data on criterion-related validity, we were greatly disturbed to find the psychologists enthusiastically promoting the uncritical acceptance and use of the data. We found

the psychologists making statements such as "...we know in many instances the precise mathematical manner in which

we don't do this, that I'm aware of. We encourage them to use it as additional information, not side-source basis for selection or rejection.

1/ Standards for Educational and Psychological Tests. Washington, D.C.: American Psychological Association, 1974.

2/ Op. Cit. p. 38304-5.

ADMINISTRATIVE-INTERNAL USE ONLY

~~ADMINISTRATIVE-INTERNAL USE ONLY~~

selected test scores correlate with performance in a number ^{1/} of specific job settings." Even under the most ideal conditions for doing criterion-related studies one never knows this precisely; one only has estimates of the relationship. The estimates that the psychologists have, from the studies that we examined, are unreliable and unstable. We think the psychologists should show more restraint both in their statements about validity and their use of the validity data.

Reliability of PATB

Little attention has been given to determining the reliabilities of the scores from PATB. Since 1958, only one reliability study has been done. This study and the Test Data Book No 15, 1 July 1958 are the only sources of information on reliability.

No reliability data are available for minorities or for females on the work attitude scales. Until reliability

1/ Memorandum from C/PSS to DDA, 25 July 1979.

~~ADMINISTRATIVE-INTERNAL USE ONLY~~

ADMINISTRATIVE-INTERNAL USE ONLY

data are available, these tests should not be used to make decisions about these individuals.

Reliability, i.e., the accuracy or stability of scores, is an important characteristic of a test. If the scores on a test are unreliable, then the scores will not predict anything. For making decisions about individuals one needs to have reliabilities in the high .70's or .80's. For white males and females, the scores on Reading Vocabulary, Reading Comprehension, Arithmetic Problems, Numerical Operations, and Interpretation of Data tests have at least marginally acceptable reliabilities. The reliabilities of the Figure Matrices, Contemporary Affairs, and Considerations tests are unacceptably low. The reliabilities of the Work Attitude scales for white males and the Temperament scales for white males and females are also too low to be used to make decisions about individuals.

we are using these in a descriptive manner, not a predictive in our test write-ups. !!!!!

There are no reliability data for the writing sample.

This test needs to have two kinds of reliability established, the reliability of the sample as representing the true writing ability of the applicant and the reliability of scoring or judging the quality of the writing. Neither type of reliability has been determined.

→ This writing sample is the first (and sometimes only) indication we have of an applicant's writing skill. Samples submitted by applicants ²⁵ often have been edited (as

ADMINISTRATIVE-INTERNAL USE ONLY

in term paper and are sometimes ghostwritten by the authorship.
Approved For Release 2002/01/25 : CIA-RDP00-01458R000100130008-0
(agency has spent much \$ on typists than people to write)

~~ADMINISTRATIVE-INTERNAL USE ONLY~~

The Narrative Report

Unit managers never see the actual test scores of candidates. They have access only to the psychologist's narrative report which consists of 7 sections, 6 of which describe the candidate's performance on the tests and 1 of which makes specific recommendations about the candidate. We have given a detailed critique of the narrative report in Appendix 2. In this section we will summarize the main points in our critique.

The last section of the narrative report, Comments and/or Recommendations, caused us considerable concern. We found that the psychologists were making very strong recommendations to hire an applicant for a specific job or for a specific unit, but we found no data to support these recommendations. The psychologists stated that these were based on "test profiles" that they have developed for various Agency jobs. We asked for, but were not given, any information on how these were developed. We suspect that the "test profiles" are nothing more than the average scores made by a group employed in a certain job and that they are based on the same small samples used in the validity studies. If they are based on these samples, they are unreliable and unstable and should not be used. Even if the test profiles

~~ADMINISTRATIVE-INTERNAL USE ONLY~~

~~ADMINISTRATIVE - INTERNAL USE ONLY~~

are based on large samples, they represent nothing more than descriptions of how the employed group, on the average, performed on the tests. They do not provide evidence that applicants need to have the same profile to perform adequately on the job.

but he does have to complete something in "our society" !!!

No, but literacy is a requirement for most jobs (except possibly some of IG staff)

The psychologists also appear to be using the multiple regression equations or the discriminant analysis equations generated in the studies of validity to make recommendations. We have previously pointed out that these equations were based on small samples and have not been cross-validated and that the samples used to generate them have been composed primarily or ^{where is this data coming from?} solely of white males. Recommendations for specific types of employment made without adequate validity data, as are these that we are discussing, promote unfair use of the test results. Such recommendations tend to exclude from consideration for employment those individuals who score low on the tests when there is no evidence to indicate that these people could not perform satisfactorily on the job. This practice violates EEOC guidelines.^{1/}

WRONG

The descriptive sections of the narrative report also caused us considerable concern. We have previously pointed out that there is no evidence on the content or construct validity of the tests in PATB; therefore the psychologists'

1/ Op. Cit. p. 38301

~~ADMINISTRATIVE - INTERNAL USE ONLY~~

ADMINISTRATIVE-INTERNAL USE ONLY

inferences about what the tests are measuring cannot or should not be made. We also found that the psychologists were reporting scores on the tests without regard for the varying reliabilities of the tests. We found that the psychologists were rather consistently misinterpreting the results of the Strong-Campbell Interest Inventory by inferring abilities or personality characteristics from the scores on these tests, which cannot be validly done. We found that the reports of writing ability varied so extensively from one report to another that it would be difficult for the user of the report to compare the writing ability of different applicants.

not so! We also have abilities scores, not justSCII data.

This is a non-sequitur. Reports of writing ability vary, probably because writing ability itself varies from case to case.

No systematic studies of the narrative report have been done. There are no guides for the psychologists to use in preparing the narrative reports. We suspect, but cannot prove, that the manner in which the narrative report is written greatly influences the managers in the units who are using the reports to make hiring decisions. The lack of uniformity in describing test performance in the narrative reports raises questions as to whether applicants scoring at the same level on the tests are perceived in the same way by persons in the unit who are making the employment decisions.

This is why we monitor the work & coach new staff members

When the writing is especially poor, we often send a copy of the essay to ²⁸ the requester.

ADMINISTRATIVE-INTERNAL USE ONLY

In view of the above, we question whether the present form of narrative reporting represents the most desirable form for making test scores available to potential employers in the Agency. Preparation of reports is costly in time of the psychologists and of typists. To the extent that the narratives are not standardized and stereotyped, they tend to branch out into interpretations that are subjective and, many times, invalid. We would propose the uniform preparation for each applicant, at least for the cognitive tests in the battery, of (a) a profile report of scores or (b) a completely standardized verbal report. Either of these could be generated by a computer.

It is not the most desirable form; we do need to go to a computer-generated reporting system.

much more difficult to measure - than interpretation by typists

Fairness or Bias in PATB

The terms, test fairness or test bias, have no uniform definition. One definition of test bias is based on lower average performance of certain groups on the tests; however, this definition is completely inadequate. According to this definition, a spelling test would be biased against those who cannot spell and reading tests would be biased against illiterates. For the purposes of this report, we will use the definition of unfairness given in the EEOC guidelines ^{1/} which is as follows:

1/ Op. Cit. p. 38301

ADMINISTRATIVE-INTERNAL USE ONLY

When members of one race, sex or ethnic group characteristically obtain lower scores on a selection procedure than members of another group, and the differences in scores are not reflected in differences in a measure of job performance, use of the selection procedure may unfairly deny opportunities to members of the group that obtains the lower scores.

1/

The EEOC guidelines also state that organizations using selection procedures should determine whether these have "adverse impact" defined as a selection rate for any race, sex or ethnic group which is less than four-fifths (eighty percent) of the rate for the group with the highest rate of acceptance. We searched for but could find no evidence that studies of adverse impact of PATB or any of the other procedures used to select personnel for Agency jobs have ever been done.

Did they deliberately ignore the study JHW sent?

Interesting, in light of below

We found only two studies of minority applicants. One of these done by [REDACTED] in 1973 has to be disregarded because the data were not analyzed correctly. The other done by the staff of PSS on black applicants from 1 January 1974 through 10 January 1977 showed that of 958 black applicants, 438 did not take PATB and 520 did take it. Out

oops!

1/ Op. Cit. pp. 38297-8.

ADMINISTRATIVE-INTERNAL USE ONLY

~~ADMINISTRATIVE-INTERNAL USE ONLY~~

of this number of black applicants, 15% were hired by the Agency. Of those hired, 57% were tested by PATB and 43% were not tested. It is difficult to interpret these data in terms of fairness of PATB. About the only thing that one can say about them is that PATB does not appear to have any more bias than do the other selection procedures.

Hmmm

There are, though, a number of factors that indicate that there is potential for unfair use of PATB. These are:

1. The tests were normed on largely white groups.
2. The samples used to determine criterion-related validity have been composed largely or solely of white males.
3. The equations that are being used by the psychologists to make recommendations about hiring an applicant are based on small samples of white males and have not been cross-validated.

Additional Comments

Despite the fact that we have been very critical of PATB and the validity evidence for it, we think that the Agency needs a battery of good selection tests for the following reasons. First, if the tests are eliminated, the only procedures for selecting personnel would be interviews, review of past academic records and experience, and letters

~~ADMINISTRATIVE-INTERNAL USE ONLY~~

of recommendation. The extensive research literature on the use of interviews for selecting personnel shows them to be invalid, unreliable and subject to the personal biases of the interviewer. Letters of recommendation generally give little useful information about applicants. Past academic records are difficult to interpret because of grade inflation and because of differences in characteristics of the student bodies (and in grading practices) in different colleges. If, in interpreting past academic records, one gives too much weight to the selectivity of the institution attended, one is likely to belittle high level achievement in a non-selective institution. Yet many very able people graduate from non-selective institutions.

Second, the number of applicants for professional jobs in the Agency is very large in relation to the number of jobs to be filled. It would be impossible to interview intensively all applicants for jobs in the Agency. A valid and reliable test battery could help to identify those applicants who are most promising for jobs in the Agency.

Third, from reading such job descriptions as were available, and from talking with Agency personnel, it appears that many of the jobs do make demands for a high level of ability on cognitive functions of obtaining,

~~ADMINISTRATIVE-INTERNAL USE ONLY~~

synthesizing and processing information. These include such things as reading a wide variety of documents, which may include quantitative, tabular or graphic content as well as prose; evaluating the reliability and relevance of statements from individuals and from documents; combining incomplete and sometimes inconsistent or contradictory information from various sources into a coherent synthesis; and preparing clear, concise reports that summarize the information and propose conclusions to be drawn from them. All of these abilities can be measured by paper-and-pencil tests, although the present battery does not appraise them very well.

*only not ?
obvious and
immediate
suggestions?*

The present battery of tests does the best job of providing information about abilities to comprehend the information input. This is gotten at through tests of vocabulary, reading comprehension, and arithmetic reasoning. However, the other abilities are not adequately appraised.

① You don't know this.

② The topic of the sample has little to do with our ability to evaluate the individual's writing skills.

A writing sample is obtained, but it does not seem to be closely related to the writing that employees will be called upon to do in their job. We would suggest that consideration be given to replacing the present writing sample with a standard task in precis writing, in which

~~ADMINISTRATIVE-INTERNAL USE ONLY~~

~~ADMINISTRATIVE INTERNAL USE ONLY~~

the applicant would be given a body of information in a mixed up order, some relevant and some not relevant, and would be required to produce a coherent, brief (perhaps 250 words) summary of the material. Another task that might be tried as a job related writing performance task would involve selection, evaluation and synthesis of information from several partially relevant, incomplete and somewhat contradictory sources. Both of these could be scored more reliably than the present writing sample and both would be more job relevant.

*I doubt it.
irrelevant to the
purpose of a
writing sample*

We have serious doubts or reservations about the usefulness of the other tests in PATB. The Interpretation of Data test has marginal reliability and dated content.

WHAT? I thought they said there were no data on this.??

The Numerical Operations test which appraises speed of simple numerical computation seems of somewhat doubtful

*Perfect example
of a Schoolteacher's
view. No applied
experience shows
through in this
kind of statement.*

relevance in view of the nature of Agency jobs and of the universal availability of hand calculators.

The Considerations test has a very low reliability and is appraising a type of verbal fluency that does not appear to be relevant to the performance of professional jobs in the Agency.

Where does this data come from?

Schoolteacher

We question, both on psychometric and on policy grounds, the use of the self-report instruments that allegedly appraise work attitudes and temperament. They have

*34 → it is related to ops work
- educational fluency
- creativity etc*

~~ADMINISTRATIVE INTERNAL USE ONLY~~

ADMINISTRATIVE-INTERNAL USE ONLY

shown no consistent correlations with job performance.

Their reliabilities are unacceptably low. They have no demonstrated construct validity, require the individual to

"testify" against himself/herself, and are subject to

faking. *→ If a test profile (Work Attitudes) looks too good, we look for B+ item to support it. If the profile shows inflexibility, we accept that as a true statement*

We have very strong objections to the use of the

Strong-Campbell Interest Inventory for selection of personnel. It lacks Agency norms and Agency validation. It is consistently misinterpreted in the narrative reports.

Research on the Strong-Campbell has shown that it has validity only for entrance into and persistence in an occupational field. Scores on the instrument have not been shown to relate to successful job performance.

There are other aspects of the tests that indicate the need for revision. The content of some of the tests is out of date; e.g., prices appearing in some arithmetic problems. The Reading Comprehension test requires the applicant to answer 41 items but only 31 are scored. The passages in that test are predominantly drawn from literary essay type materials that tend to focus on middle-class manners and conventions. The Vocabulary test contains

ADMINISTRATIVE-INTERNAL USE ONLY

~~ADMINISTRATIVE-INTERNAL USE ONLY~~

too many esoteric words such as scantling and mignonette whose job relevance is difficult or impossible to discern. The "face validity," i.e., the appearance of job relevance, of most of the tests could be improved by making the content more related to that encountered in the work of the Agency. Although "face validity" has little effect on the real validity of tests, in today's contentious climate and controversies about tests, face validity will reduce some of the rhetoric.

accuracy

Baloney.

Although we have suggested certain revisions be made in PATB, we strongly recommend that no revisions be made until a systematic analysis of professional jobs in the Agency is done. The method of job analysis developed by McCormick (1969, 1972, 1977)^{1/} appears to be especially

1/ McCormick, Ernest J. The Development and Background of the Position Analysis Questionnaire (PAQ) Lafayette, Indiana. Occupational Research Center, Purdue University June 1969.

Marquardt, Lloyd D. and E. J. McCormick. Attribute Ratings and Profiles of the Job Elements of the Position Analysis Questionnaire (PAQ). Lafayette, Indiana. Department of Psychological Services, Purdue University. 1972.

McCormick, Ernest J., A. S. DeNisi, and J. B. Shaw. The Use of the Position Analyses Questionnaire (PAQ) for Establishing the Job Component Validity of Tests. Lafayette, Indiana. Department of Psychological Services, Purdue University. 1977.

~~ADMINISTRATIVE-INTERNAL USE ONLY~~

ADMINISTRATIVE-INTERNAL USE ONLY

*This would have
to be done at a
high level. PSS does
not have the clout
to impose job analyses
on user offices.*

suitable for Agency type jobs. The job analyses will provide information concerning the knowledges, skills and competencies needed to perform professional jobs in the Agency. One will also be able to determine whether these knowledges, skills and competencies are common across all or most professional jobs in the Agency. If they are, then one battery of tests is needed. If they are not, then more than one battery is needed. Once the skills and competencies needed to perform professional jobs in the Agency are known, then one can construct tests or devise other procedures to appraise them. To undertake the construction of a new battery of tests or the revision of the present battery without a systematic job analysis is likely to result in no improvement.

We were somewhat distressed to find that the results from the PATB, especially those from the cognitive tests, were made available only as a secondary source of information, and then only when specifically requested by a potential employer. No test results are available with the preliminary prospectus about a candidate that is circulated to the various branches within the Agency. For a great many applicants, the PATB test scores never

*managers
should be given
at least the
opportunity to
make
independent
judgments.*

ADMINISTRATIVE-INTERNAL USE ONLY

ADMINISTRATIVE INTERNAL USE ONLY

reach a potential employer. When this fact is combined with the finding that 80% of those in the highest group on the Intellectual Composite are not employed by the Agency, one wonders whether optimum use is being made of the test results. Obviously, there are many other reasons why these capable applicants are not employed, but it would seem that a good deal of talent is "blooming unseen" because the information does not reach potential users.

*of course
these are -
by necessity, needed
then having undown
etc*

We would like to see some type of procedure set up through which each of the more able applicants would at least be sure to be considered by potential employers.

We are not sure what would be the best operational procedure. One procedure would be to prepare a weekly (?) list giving the names of high scoring candidates, and to circulate this routinely to potential employers. The employer could then request the full personnel file for cases in which there was interest. Another procedure would be to routinely send the computerized test report for all cases to the Skills Bank where (1) it would become part of the individual's file, and/or (2) salient points would be incorporated in the precis on each individual that is circulated to potential employers.

*What the hell
does this
mean for EEO,
the focus of
most of their
criticism?*

*Talk about unfair!
This has got to be illegal, too!*

ADMINISTRATIVE INTERNAL USE ONLY

ADMINISTRATIVE-INTERNAL USE ONLY

We are not well enough versed in internal Agency operations to determine what procedures would be best to ensure that all the most capable applicants receive serious consideration. We do believe that the test results can and should have a more active and positive role in employment decisions.

*we haven't
thought
no*

*True!
what would
success system do
to for
reliability
studies
they never
about!*

Another aspect of the recruitment process that aroused our concern was the relatively short time that a candidate's file remains in the active Skills Bank file, and the relative inaccessibility of that record once it is removed from the active file. We understand that procedures are under way to computerize the files of applicants, and that it will become feasible to store a large pool of cases, coded by various relevant facts about them, so that cases showing certain types of experience or certain skills can be readily retrieved. We consider this a very important constructive step. With the sizeable investment in recruiting and testing, it seems most unfortunate that capable individuals be lost just because there is no vacancy that happens to fit their special competencies just at the time when their application is being processed. A computerized record retrieval system should make it possible to exploit more fully the talents located in the recruiting effort.

*I could do
this, but it
would take a great
deal of cooperation from
ODP and OP to make
it work.*

ADMINISTRATIVE-INTERNAL USE ONLY

We would urge that the retrieval system include ^{not for dissemination in raw form} information on test scores as well as on relevant background and skills. It would be helpful to be able to sort out, for example, not only all political scientists who read Russian, but also the sub-group of those political scientists who also analyze data effectively and write clearly and concisely. The person(s) so identified might no longer be available for employment when a vacancy arose, but it would be nice to be able to identify such persons and explore further their desirability and availability.

I completely agree with their conclusion here.

Once again, we do not have the background to suggest detailed operational procedures. However, if the desirability of the goal is accepted, we feel confident that procedures could be worked out.

Major point: they totally disregard information (and clinical verification) of PTB data gathered in literally thousands of cases where we counsel the employee for vocational purposes, where we seek verification of hypotheses generated from test patterns through interview questions, where we "trouble shoot" problem cases with mgmt. etc. They have missed a very major point in their conclusion & recommendations. Our staff, after training and coaching, uses the test battery as a clinical ⁴⁰ as well as an actuarial instrument.